Transparent_Food

Contract No.: FP7-KBBE-2009-245003
Transparent_Food
Quality and integrity in food: a challenge for chain communication and transparency research

Coordination and Support Action – CSA

Food Quality and Safety

D 2.2: Feasibility study taking into account different ways of serving tracking/tracing needs within the sector

Due date of deliverable: Project month 12 (November 30, 2010)
Actual submission date: January 31, 2011

Start date of project: 01 December 2009 Duration: 24 months

Lead Contractor for this Deliverable: KTBL, Germany

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	
PP	Restricted to other programme participants (including the Commission Services)	**X**
RE	Restricted to a group specified by the consortium	
CO	Confidential, only for members of the consortium	

D 2.2: Feasibility study taking into account different ways of serving tracking/tracing needs within the sector

Deliverable datasheet	
Project acronym	Transparent_Food
Project full title	Quality and integrity in food: a challenge for chain communication and transparency research
Project contract No.	FP7-KBBE-2009-245003
Dissemination level	PP
Official delivery date	30 November 2010
Organisation name of lead contractor for this deliverable	KTBL, Germany
Authors	Esther Mietzsch (KTBL) Daniel Martini (KTBL) Dr. Martin Kunisch (KTBL)
Editing	KTBL, Germany
Version and date	V1.0 / 31 January 2011
Approved by	Prof. Dr. Gerhard Schiefer

Contents

Executive Summary

One of the objectives of the Transparent Food project is to create a blueprint proposal for a European Backbone Solution that provides basic and simple functionalities to enable integration of different systems for tracking and tracing in the food sector as a prerequisite to transparency.

This document provides an overview report on the conducted analysis and evaluation of technologies that can be used to build such a solution. Four aspects have been taken into account for now: Identification mechanisms and methods to generate unique identifiers and implement labeling, protocols providing a mechanism of data exchange, syntax as a means of structuring information and semantics giving meaning to data items and information sets. Experiences from other sectors have been collected as well. This document can thus serve as a feasibility study giving information on capabilities of certain technologies. Based on this work and the requirements analysis, of which the results are available in D2.1, the project will continue onwards to draft the backbone solution specification proposal.

The document collection – mainly consisting of specifications and technical design and architecture descriptions –, for which this report is a summary, currently still continues to grow. This report is thus expected to be revised during the further course of the project.

List of Figures

List of Tables

List of Appendixes

List of Abbreviations

EAN	European Article Number
EDI	Electronic Data Interchange
EFSA	European Food Safety Authority
EPCIS	Electronic Product Code Information Services
EPC	Electronic Product Code
FTP	File Transfer Protocol
GTIN	Global Trade Item Number
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
IETF	Internet Engineering Task Force
ISO	International Standards Organization
OASIS	Organization for the Advancement of Structured Information Standards
RDF	Resource Description Framework
REST	Representational State Transfer
RFID	Radio Frequency Identification
RPC	Remote Procedure Calls
SKOS	Simple Knowledge Organization System
SMTP	Simple Mail Transfer Protocol
SSL	Secure Sockets Layer
TLS	Transport Layer Security
TRU	Traceability Reference Unit
UBL	Universal Business Language
UNECE	United Nations Economic Commission for Europe
UN/CEFACT	United Nations Centre for Trade Facilitation and Electronic Business
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
VAN	Value Added Network
W3C	World Wide Web Consortium
WSDL	Web Services Description Language
XML	eXtensible Markup Language

1. Introduction

To achieve appropriate transparency across the whole food sector, different tracking and tracing systems have to be interconnected. There is currently no satisfying solution available that is up to this task. One of the objectives of the Transparent Food project is thus to create a blueprint proposal for a European Backbone Solution that provides basic and simple functionalities to enable integration of different systems.

As a first step, a requirements analysis for this backbone solution has been conducted. Within this analysis, a number of challenges have been found.

There are e. g. scalability requirements of the solution resulting from the sector structure. Methods and technologies used will have to accommodate on the one hand large amounts of smaller data packages and on the other hand a large number of small stakeholders. Data from a chain are used by different stakeholders in different scopes (e. g. organic production only, cold-chain-management only etc.). The demands upon providing the possibility to allow for different views and for a number of varying analysis methods are high. It does not suffice to provide tracking and tracing only to achieve transparency in the food sector. A number of other attributes are of increasing interest, not only to allow for quality control but to also be able to market quality products. To accommodate this demand, it is important that the data models used are designed to allow for extensibility without breaking backwards compatibility from the ground up. It is a crucial factor for success to find a set of standards and methods that are up to the task but at the same time simple, clear and generic enough to be accepted by system providers with their diverse development and computing environments and toolsets. Ideally, a set of orthogonal specifications without unnecessary duplication of functionalities is chosen. The task of building a backbone can be facilitated to a great extent by reuse of existing standards and leveraging the capabilities and networks of existing organizations. The document will thus start out with an overview on organizations which create open standards, data models and protocols for use in supply chain management.

It continues providing a report on work conducted on assessment of existing and emerging technologies as to their capabilities with regard to the requirements found in the requirements analysis. Four aspects have been identified that have to be considered on the technical level in building a backbone infrastructure: identification of items, protocols used in communication, syntax and semantics of data exchanged. Technologies have been classified according to these aspects. A short discussion on what role the respective technologies can play in a food sector tracking and tracing backbone is given. The specifications of the standards mentioned are available, provide more detail and are listed in the references appendix

It turned out, that there are other sectors with problems similar to the ones to be found in the food sector. Another part of the document thus focuses on systems and common practices in use in other sectors and evaluates these with regard to their applicability in the food sector.

2. Organizational Resources

Various organizations provide standards or directories for data in supply chain management and in the food and agricultural sector. Both public and governmental bodies and private associations are involved in this work.

The GS1 is a global non-profit association with over 100 member organisations. GS1 has developed standards for identification such as key numbers (e.g. GTIN), data carriers (bar codes, EPC) and communication (e.g. EANCOM). The development process (Global Standards Management Process) involves working groups with experts delegates from member companies. The responsibility of the national member organisations is the allocation of unique numbers to member companies and providing training and support. The member companies have to pay a fee for obtaining the GS1 company prefix.

A subsidy of GS1 is EPCglobal. Its subscribers are end-users such as manufacturers and solution providers such as hard- and software companies. EPCglobal is leading the development of industry-driven standards for the Electronic Product Code (EPC) to support the use of Radio Frequency Identification (RFID). It assigns EPC Managers Numbers, delivers certification of application and provides other services. Member companies can participate in the development of standards.

Another not-for-profit organisation is OASIS (Organization for the Advancement of Structured Information Standards). Members are government agencies, software providers and industry groups. Standards developed by OASIS are for example the Open Document Format for Office Applications and ebXML. The specifications are all royalty-free.

Also involved in the development of ebXML is the UN/CEFACT (United Nations Centre for Trade Facilitation and Electronic Business). It also produced the UN/EDIFACT standard. UN/CEFACT provides recommendations for trade facilitations, electronic business standards and technical specifications. Members are governments, intergovernmental organisations, NGOs, chambers of commerce, and companies from the private sector. The permanent groups which develop standards, specifications and guidelines are open to technical experts in the respective fields.

The UN/CEFACT has been established by the UNECE, the United nations Economic Commission for Europe. The major aim of the UNECE is to promote pan-European economic integration. The Working Party on Agricultural Quality Standards developed a series of standards and recommendations on food products such as meat.

On EU level, several organisations are involved in the standardisation process. The European Food Safety Authority (EFSA) operates separately from the European Commission, European Parliament and EU Member States and is governed by an independent Management Board, but is funded by the EU budget. The EFSA collects analytical measurement data for the presence of harmful or beneficial chemical substances in food and feed from a variety of providers such as national authorities, laboratories, research institutes etc. A standard sample description for food and is used. The standard sample description document was developed by the Technical Working Group on Data Collection (TWG-DC).

Other efforts come from limited term projects. EuroFIR (European Food Information Resource) was a five-year Network of Excellence funded by the European Commission's Research Directorate General under the "Food Quality and Safety Priority" of the Sixth

Framework Programme for Research and Technological Development. It contributed to the cataloguing of food products.

On a national level, various governmental bodies such as ministries, agencies, federal state offices produce standards for data exchange in the food sector. In Germany for example, the governmental agency for consumer protection and food safety (BVL) is coordinating the creation of coding catalogues and mechanisms for data exchange for food control purposes.

Apart from those stakeholders providing domain specific standards, there are a number of organizations involved in providing basic and generic information technology standards. The most notable ones with this regard include the World Wide Web Consortium (W3C) offering a multitude of recommendations for data and information exchange across the Internet directed to humans and machines, the Internet Engineering Task Force (IETF) focusing more on the lower level protocols and machine-machine interaction or the International Standards Organization (ISO) Joint Technical Committee 1 (JTC1) serving as a normalization organization in information technology.

3. Protocol

3.1. eBusiness, Supply Chain and Food Sector Data Exchange Protocols

Electronic data interchange (EDI) enables companies to exchange business documents in a standard format. One standard for EDI is the UN/EDIFACT (United Nations Electronic Data Interchange for Administration, Commerce and Transport), which has been developed by the UN/CEFACT (United Nations Centre for Trade Facilitation and Electronic Business) since the eighties. A subset of this standard is EANCOM, which has less optional elements and is easier to handle. The EDIFACT standard lists more than 200 message types, each with a six character name (e.g. ORDER for purchase order message). A message has a hierarchical structure and is a collection of segments, which are characterized by a three character tag and conditional or mandatory data elements. Single characters are used as field separators and terminators. Multiple messages can be grouped together in an interchange and are wrapped into an electronic envelope also consisting of segments. The syntax of EDIFACT is very condensed and not meant to be human readable.

EDIFACT messages can be sent from one company to another using any available communication protocol. In the beginning of EDI, dedicated lines or modems where commonly used. Another way of message exchange are Value Added Networks (VAN) realized by provider companies which simply act as an electronic mail box. The internet protocols (SMTP, HTTP(S), FTP) are also used. Based on HTTP is the specification AS2, which uses signing, encryption and MDN (Message Disposition Notification, the ability to provide return receipts). It is widely used in the retail sector. For smaller enterprises without own EDI infrastructure who only have to transmit smaller amounts of data, web based applications (WebEDI) are available.

A newer standard for electronic business is ebXML (Electronic Business using eXtensible Markup Language). It is maintained by the UN/CEFACT and by OASIS (Organization for the Advancement of Structured Information Standards). The first version was already issued in May 2001, since then a number of its specifications have become ISO standards. ebXML includes five types of specifications: on business processes (Dubray et al., 2006), on collaboration protocols and agreements (OASIS ebXML Collaboration Protocol Profile and Agreement TC, 2002), on messaging services (ebMS; Wenzel, 2007) on registries and repositories (Fuger et al. 2005a and 2005b) and on core data components. All definitions of the data exchanged over ebXML are stored in an ebXML registry as XML documents. The data pools are managed by service providers or major suppliers. ebMS is based on SOAP (Mitra et al., 2007), the underlying communication protocol is usually HTTP. SOAP Version 1.2 is a lightweight protocol intended for exchanging structured information in a decentralized, distributed environment. "Part 1: Messaging Framework" defines, using XML technologies, an extensible messaging framework containing a message construct that can be exchanged over a variety of underlying protocols.

EPCIS (Electronic Product Code Information Service; EPCglobal Inc., 2007) is a standard for the exchange of data on product movement. It is meant to be complementary to EDI. Each time an electronic product code is read, an event is generated which answers the questions about what (product identified by manufacturing data e.g. EPC number), where (location of enterprise, position in supply chain), when (time of event) and why (status, process step). The events are stored in a database, the EPCIS repository, usually hosted by the manufacturer. The repository has a capture interface for storing the events and a query interface for retrieval of event data. An ONS (Object Name Service) provides a lookup service which delivers the address (URL) of an EPCIS repository for a given EPC. Other EPC Discovery services which serve as a search engine for EPCIS are currently under development. The transfer of data via the capture interface is via HTTP, the query interface uses SOAP, XML over AS2 and XML over HTTP(S). All message protocols must be able to use authentication and authorization.

A complete free and open source implementation of the EPCIS standard specification including Repository as well as Query and Capture clients has been developed by a team located mainly at the ETH Zurich/Switzerland.

A detailed guidance document on data exchange has been published by the EFSA in 2010 for their purposes (EFSA, 2010b). The main requirement was the simplicity of the protocol and its easy implementation. The transmission of data might be either by manual posting of files (upload to a web application) or by automatic transmission. As file formats, Microsoft Excel files or CSV files (comma separated values) are permitted for a limited period only, as these formats are more susceptible to errors and automatic validation is more demanding. The preferred format for the data is XML.

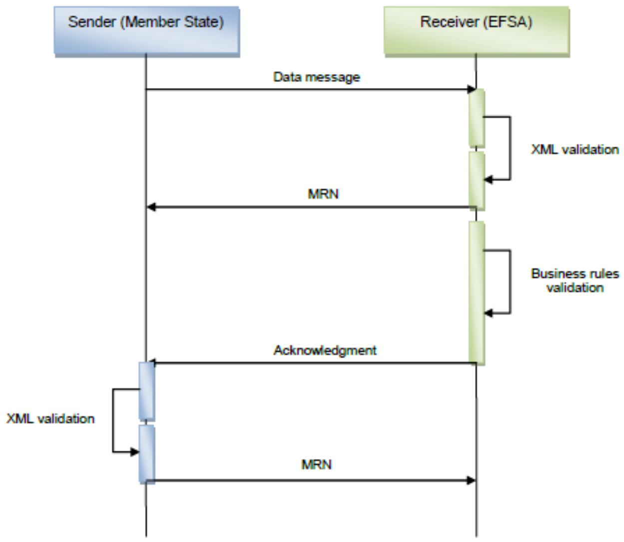

Fig. 1: EFSA Message exchange protocol

The Message Exchange Protocol describes the exchange of messages between sender (e.g. member state authorities) and receiver (the EFSA): the data message, the MRN message (Message Receipt Notification) and the acknowledgement message. The transport layer for the physical exchange of the messages can be FTP or by web services. To meet security requirements, the sender software has to provide an user identification and password and to use a secure internet protocol such as FTP through SSL.

3.2. Considerations for the Tracking and Tracing Backbone

In a large and uncontrolled network, designing a robust protocol can quickly become a challenge. Using the paradigm of message orientation prevalent in networked systems in the 80s and 90s of the previous century in such a setting often leads to an unmaintainable mess of messages and method calls that have to be implemented by each stakeholder individually to be able to interoperate. Without knowledge of the exact implementation on the other end, no communication is possible. Criticism on EDI and EDIFACT mostly circles around this issue. The World Wide Web Consortium's standards around message oriented web services like SOAP (Mitra et al., 2007; Gudgin et al., 2007a and 2007b) and WSDL (Chinnici et al., 2007a and 2007b) are nowadays seen as a failure in large scale deployment for this reason also by many prominent researchers and developers in the area (see e. g. http://www.infoworld.com/d/developer-world/sun-technologist-soap-stack-failure-415?source=fssr).
Concerning the tracking and tracing backbone envisioned, focusing on these technologies can quickly lead to a very complex undertaking considering the number of stakeholders involved and the dynamic nature of the food network. For that reason, it is proposed to instead follow a RESTful paradigm (Fielding, 2000; Richardson et al., 2007), in which method calls are restricted to the minimum set that is necessary to retrieve and post data. Nature and content of the data is of no relevance on the protocol layer. Only if the data is to be interpreted later on, knowledge of semantics is required. Experiences

from system providers present in the working group indicate that this approach is feasible.

4. Syntax

4.1. Syntax Specification Mechanisms

When exchanging data, both communication partners (sender and receiver) have to agree not only on a protocol for data exchange, but also on syntax. The syntax describes the structure of the data. Technically, the data fields and how they can be identified and separated out ("tokenized") for further processing have to be defined.

The most common language for data structuring is currently XML (eXtensible Markup Language; Bray et al., 2008). XML files are text files, where each data field is marked with a tag pair. Hierarchical, tree-like structures can be implemented by nesting the tags accordingly. Crosslinks among tree elements are possible using either the ID/IDREF mechanism (within the same document) or XLink (across different documents; DeRose et al., 2010). XML files can accommodate almost every language of the world by using one of the Unicode Universal Character sets, most commonly UTF-8. As described below, XML is used in major supply chain, eBussiness and Food Sector standards. A variety of tools and libraries for almost every programming language is available. Most of the system providers analyzed in deliverable 2.1 of this work package explicitly state in their documentation that they support either XML or standards based on XML like EPCIS. In general, XML is thus a natural choice for implementing at least a part of the backbone solution components. Data structures can be specified using so-called schema languages like XML Schema (Fallside et al., 2004; Thompson et al., 2004; Biron et al., 2004) or RelaxNG (Clark et al., 2001). That way fields that may be contained in documents and their relation to each other can be restricted to a valid set that can be checked upon reading or writing files.

XML has received criticism for being too verbose and blowing up simple and small datasets and thus produces relatively large files. Therefore, alternatives have been devised, that allow for a more compact representation and quicker reading, parsing and serialization. Recently, JSON (Crockford, 2006) catches increasing attention, as it allows for easy integration of web browser application frontends into an infrastructure of services targeted at machine to machine communication. XML files can easily be converted into JSON syntax using freely available tools. A JSON schema implementation similar in functionality to XML schema has been proposed (Zyp, 2010), it is however not yet an accepted IETF standard. JSON is well suited to the requirements of transporting a large number of small data packages and of creating tracking and tracing helper components for commodity hardware that arise in implementing the backbone solution. In fact, already a number of system providers can offer support for JSON syntax.

4.2. XML in eBusiness, Supply Chains, Food and Agriculture

For ebXML, an implementation of the Core Components Technical Specification (UN/CEFACT, 2003) is the Universal Business Language (UBL; Bosak et al. 2006). This standard data format has been defined by OASIS and provides XML schemas for business documents (e.g. order, invoice, etc.).

EPCIS (EPCglobal Inc., 2007) also uses XML as data format. The messages used are centered around the concept of an "event". To describe events, four event types exist, which share some common attributes, but also have specific data fields. The types are Object Event (events during the lifetime of an object), Aggregation Event (describing the physical aggregation of children objects to a parent, e.g. cases to pallet), Quantity Object and Transaction Object. Figure 2 gives an overview of the underlying data model.

In agriculture, agroXML is striving for creating an XML-based data format for documentation in quality assurance in agriculture. The focus is much broader than only tracking and tracing. It can serve to record accompanying data of most of the processes and practices in arable farming, like soil tillage, fertilization and pest control. Subsets can be of use in the tracking and tracing backbone. For example, a common issue are changes in field geometry due to reorganization. Therefore, if track and trace is to be implemented on the field level, the traceability reference units (TRUs) to be regarded in a system change as well. Documenting these changes is possible using agroXML.

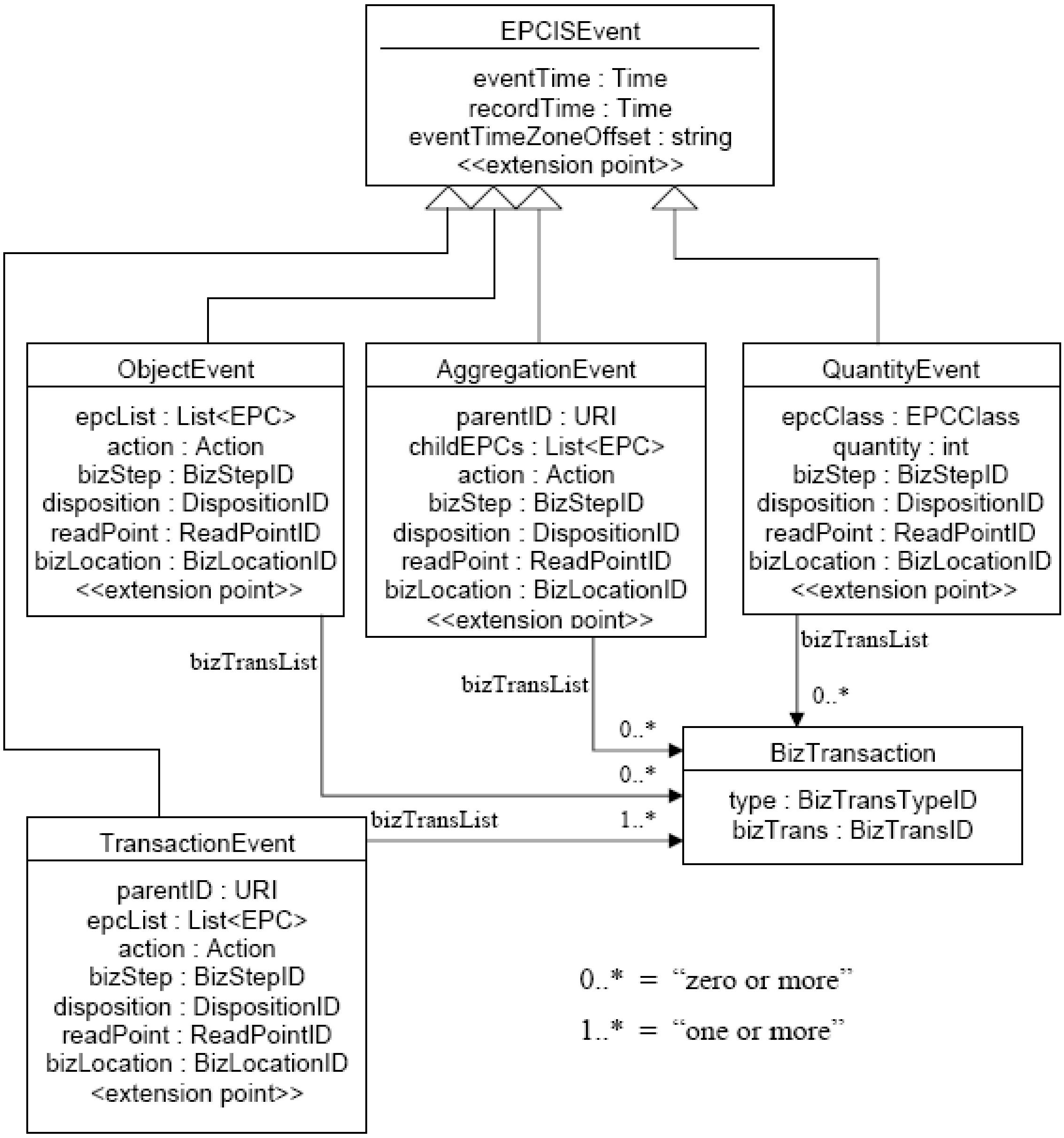

Fig. 2: EPCIS Event data model.

5. Semantics

5.1. Formalized Semantics in General

The use of controlled vocabularies is one of the issues in data exchange. Within these, a restricted number of allowed terms is defined. This avoids confusion and allows for the grouping of data and the comparability of data from different origins. Each term might have allocated a key number consisting of digits and in some cases also of letters, such as "E121".

In the past, stakeholders producing such vocabularies mostly used their own "home-grown" systems of keeping terms and relations, often using relational database technology. The drawback of this approach was that vocabularies were difficult to exchange and use across different domains and systems. Only recently, with the advent of the semantic web and a higher demand for global information exchange standards for vocabularies have gained increasing attention. Notable technologies in use include the Resource Description Framework (RDF; Manola et al., 2004; Klyne et al., 2004; Hayes et al., 2004), RDF Schema (Brickley et al., 2004) and the Simple Knowledge Organization System (SKOS; Miles et al., 2009) by the World Wide Web Consortium. They provide methods to describe terms and relations among them and to build statements describing certain resources (objects, documents, processes etc.). A URI (Berners-Lee et al., 1998) is assigned to each term and relation thus providing globally unique identification of concepts. Based upon these vocabularies, data and information can be encoded in a flexible and extensible way. Data missing in a set can often be derived using formal logics and information given elsewhere so that a networked knowledge base can be implemented. This approach would lend itself well to the backpack of information attached to a track and trace base data set described in deliverable 2.1. For this to work out, closer alignment of a number of existing vocabularies in the food and agricultural sector is required.

5.2. Domain Specific Vocabularies in the Agricultural and Food Sector

In the food sector, several classification systems have been established. Two main types can be identified: hierarchical classification and facetted classification. Hierarchical classification systems have a tree like structure where each term has a "belongs to" relationship to a parent term with a broader meaning, e.g. "apple" belongs to "fruit", "fruit" belongs to "plant products".

One of the institutions that uses a hierarchical classification system is the EFSA, the European Food Safety Authority. The EFSA collects data from the EU member states, the European Commission, the industry etc. on food consumption, the incidence and prevalence of biological risks, and occurrence of contaminants and chemical residues. A standard sample description for food and feed is used, which is composed of a list of standardized data elements (definition and structure), controlled terminologies and validation rules (EFSA, 2010). The target is to harmonize the collection of analytical measurement data on food and feed. Controlled terminologies for all parameters of the analysis have been established, e.g. for the analytical method, the country of origin, the result etc. The product code describing the product under analysis is a hierarchical tree with 376 terms such as "Lettuce" or "Goat liver", thereof 34 root terms such as "Citrus fruit", "Baby food". Each term is coded with a 9-digit product code, e.g. "P0120110A". Another list of terms describes the processes applied to the product or any indexed ingredient.

Another hierarchical classification is used by the German Federal Office of Consumer Protection and Food Safety (Bundesamt für Verbraucherschutz und Lebensmittelsicherheit, BVL). The automatic transfer of data from the federal states to the BVL on the analysis of food for the official food monitoring uses a fixed set of codes. For the description of the food and other samples, a 6-digit code is used, for example "312107: Kiwi juice", grouped under "312100: Juice from exotic fruits". This system allows for a large number of data (7 Mil in 2007) and a fast retrieval of data with a minimal effort. It has

therefore found a high acceptance. Its origin is an the early nineties, when datasets had to be as small as possible due to limited transfer rates.
Currently a project to establish a new data exchange system at the BVL is under way. The basic principles such as the uniform coding, standardized formats, central management of the catalogues and validation checks before the release of data are kept. The new system will be based on a facetted classification. A basic list of food stuffs is supplemented by a catalogue of facets describing multiple characteristics of the food. This faceted catalogue is based on LanguaL.
A facetted classification is a multihierarchical classification, where each item is described by a number of characteristics, the "facets". LanguaL (Møller et al., 2009) is a food description thesaurus which uses a facetted classification. Each food is described by a set of standard, controlled terms chosen from facets characteristic of the nutritional and/or hygienic quality of a food, as for example the biological origin, the methods of cooking and conservation, and technological treatments.

Facet	Name	Example
Facet A	PRODUCT TYPE	SAUSAGE OR SIMILAR MEAT PRODUCT (EUROFIR) [A0798]
Facet B	FOOD SOURCE	HIPPOPOTAMUS [B2130]
Facet C	PART OF PLANT OR ANIMAL	ROOT, TUBER OR BULB, WITHOUT PEEL [C0240]
Facet E	PHYSICAL STATE, SHAPE OR FORM	DIVIDED INTO SEGMENTS OR WEDGES [E0107]
Facet F	EXTENT OF HEAT TREATMENT	HEAT-TREATED, MULTIPLE COMPONENTS, DIFFERENT DEGREES OF TREATMENT [F0023]
Facet G	COOKING METHOD	DEEP-FRIED [G0029]
Facet H	TREATMENT APPLIED	OLIGOSACCHARIDE ADDED [H0240]
Facet J	PRESERVATION METHOD	PASTEURIZED BY HEAT BEFORE FILLING [J0159]
Facet K	PACKING MEDIUM	PACKED IN GRAVY OR SAUCE, VEGETABLE [K0037]
Facet M	CONTAINER OR WRAPPING	ALUMINUM TUBE, TOP LINED WITH FOIL [M0170]
Facet N	FOOD CONTACT SURFACE	BEVERAGE CAN ENAMEL, NON-CARBONATED BEVERAGE [N0012]
Facet P	CONSUMER GROUP/DIETARY USE/LABEL CLAIM	VERY LOW SODIUM FOOD [P0153]
Facet R	GEOGRAPHIC PLACES AND REGIONS	SCOTLAND [R0224]
Facet Z	ADJUNCT CHARACTERISTICS OF FOOD	APPELLATION CONTROLEE [Z0086]

Tab. 1: Facets used in LanguaL

The facet term lists are hierarchically structured. The work on LanguaL started in the late 1970's in the USA. In recent years, the EuroFIR (European Food Information Resource), an EU funded project has indexed a large number foods. LanguaL is now multilingual with approximately 70000 terms (English, German, French etc.). The main focus is on food consumption and food composition.
The advantage of a facetted classification is a greater flexibility regarding the integration of new properties. In a hierarchical classification system, a new code has to be created when a new type of a known product appears, for example a new key number has to be created if kiwi peel appeared on the market, additionally to kiwi fruit, kiwi juice etc. In a facetted classification, only a new combination of existing terms needs to be used in this case.
Using a system of key numbers for the applied terms reduces the amount of data to be transferred and stored. This was of greater importance in former years, when transfer rates and file storage place were much smaller than they are now. It also restricts the length of the codes, which facilitates automatic data processing.
In order to streamline the flow of information throughout the supply chain and to provide a standard for use between buyer and seller in the meat industry, UNECE (United Nations Economic Commission for Europe) Working Party on Agricultural Quality Standards defined the "UNECE STANDARD Bovine Carcasses and Cuts" (UNECE, 2004). Similar standards exist for caprine, chicken, duck, llama/alpaca (UNECE, 2006), ovine, porcine and turkey meat (UNECE, 2009). Each standard gives detailed specifications to identify cutting lines including colour photographs and diagrams. Also, minimum requirements for meat are formulated. All data are coded in a 20-digit string (see Tab. 2).
The UNECE purchase specification code has been assigned the GS1 application identifier (7002) to be used in conjunction with a Global Trade Item Number (GTIN) and represented in the GS1-128 bar code symbology. This allows the UNECE code information to be included in GS1-128 bar code symbols on shipping containers along with other product information.

Data field	Category	Example	Example Code
1	Species	Bovine (Beef)	11
2	Product/cut	Tenderloin	2150
3	Not used	-	
4	Refrigeration	Deep frozen	3
5	Category	Heifer	4
6	Production system	Organic	3
7a	Feeding system	Grain fed	1
7b	Not used	-	
8	Slaughter system	Halal	3
9	Post slaughter system	Specified between buyer and seller	1
10	Fat thickness	3 mm maximum	4
11	Bovine quality	Company standards	2
12	Weight range	Not specified	0
13	Packing	Cuts – vacuum packed	5
14	Conformity assessment	Trade standard conformity assessment	2

Tab. 2: UNECE Standard Bovine Meat

Multiple other lists and controlled catalogues of food stuffs are in current use. In Germany for example there exists also the "Bundeslebensmittelschlüssel" (Federal key for food) which uses a 7-digit coding system for food consumption surveys (not for tracking purposes), and the "Stoffliste des Bundes und der Bundesländer" (List of substances for the use of the Federal Government and the federal states) which gives a list of plant products used for the legal distinction between drugs and food.
The AGROVOC thesaurus by the Food and Agricultural Organization of the United Nations (FAO) is nowadays the most comprehensive multilingual thesaurus and vocabulary for agriculture. Originally, it was devised for indexing of literature, but it is increasingly used also in facilitating knowledge sharing and exchange through electronic media and machine-readable data formats. It contains approximately 30000 so-called concepts (terms) that are at least in part available in more than 30 languages. The vocabulary is provided in standard RDF and SKOS and concepts are identified by URLs. Therefore, it is easy to reference these concepts or create mappings to other vocabularies. Apart from several agricultural ontology relations (for a complete list see http://aims.fao.org/website/Ontology-relationships/sub) AGROVOC uses common thesauri relationships like "broader term", "narrower term", "related term" etc. that are also used in LanguaL. By conversion to RDF/SKOS, LanguaL could therefore probably relatively easily be interconnected with AGROVOC thus forming a larger, networked Agrifood-Vocabulary.

6. Identification

6.1. Identification in Services

When exchanging information on a number of very similar real-world objects in information technology systems, it is crucial to identify, to what object a certain information set refers. For that purpose, unique keys are commonly used. Real world objects can be marked with these keys using sticky barcode tags, RFID chips or simple manual numbering. When it comes to tracking and tracing, i. e. objects moving along a supply chain between different enterprises or stakeholders, unique identifiers across the whole chain simplify handling significantly, as otherwise, tags would have to be replaced upon arrival at the respective following party.
To achieve this uniqueness across a number of systems in a global context, often Uniform Resource Identifiers (URIs; Berners-Lee et al., 1998) are used. An advantage of them is that their syntax is standardized and an exhaustion of unique identifiers is practically impossible. A hierarchical distribution of responsibility for certain value spaces is provided for the URI subset of Uniform Resource Locators (URLs) by the domain name system registration procedures and management practices are in place and well understood.
URIs can be connected to a dereferencing mechanism: The identifier can directly be used to retrieve information or data on a certain object. The most commonly known application of this mechanism is the retrieval of websites in a web browser using its URL. The EPC Information Services make use of this facility as well by providing standardized means of how a URI is constructed from a bar code or RFID scan. Using this URI, a

lookup can follow resulting in an information request for the respective object with the tag on it. All in all, a standardized dereferencing mechanism allows for interconnected information sets by specifying further links in documents available at a certain URI (c. f. anchors in the Hypertext Markup Language, HTML). In a tracking and tracing system this cross linking can provide a very simple and powerful mechanism to build up distributed data storage along a chain.

6.2. Tagging

For tracking and tracing purposes, identification by labelling or tagging of goods has a crucial role. Labels can be either bulk-wise, such as the EAN-code (European Article Number), batch numbers and expiry dates on various products, or on an individual basis. Individual tagging has been realized in the identification of animals, but e.g. also on pharmaceuticals for fraud detection purposes.

Identification of farm animals is mandatory in the EU for bovine animals (cattle), pigs, sheep and goats, and for equine animals (horses, donkeys, etc.). The objective is the localisation and tracing for veterinary purposes and ensuring food safety. As equine animals are of a minor importance in the food chain (< 1% in all of the European countries except for Italy), they are not considered here. For poultry, identification is not required on a single animal basis.

For pigs, identification has to be by ear tags or tattoos. A register has to be kept on each holding, and the number of animals has to be recorded during movements.

For sheep and goats, electronic tagging is mandatory for all animals born after 2009-12-31, a register is kept on each holding. Currently, movements have to be recorded for each group of animals, and a database is kept at national level. For cattle, the system consists also of similar elements: individual identification, maintaining a register on each holding (farm, market etc.), cattle-passports (recording of each individual movement), a computerised database at national level. The rules on the individual traceability of bovine animals were already laid down in 1997 as a consequence of the BSE crisis. Identification of cattle is usually by double ear tags on each animal, electronic identification is currently under development. This individual identification and recording of each movement allows tracking each animal from birth to the slaughterhouse.

Electronic identification is usually realized by RFID (Radio Frequency Identification). A passive transponder receives energy and responds with the information stored in its IC (integrated circuit). Different frequencies are in use. The LF-range (< 135 kHz) is the only one suitable for animal identification as only this signal can penetrate through living tissue. The transponders are usually applied as electronic ear tags or as bolus. The bolus is swallowed by ruminants and then stays in the rumen. Injectable microchips are also available, but they are not recommended for animals for slaughter, because the post slaughter recovery is difficult. Readers can be either stationary or handheld. The use of RFID for animal identification is standardised by two ISO norms at worldwide level (ISO 11784 and 11785). The code stored in the transponder contains the country code, a 12 digit identification code and several additional fields. Manufacturer coded transponders are also available. The transponders can only store numbers, but no alpha numeric coding. The use of numbering schemes for the identification number is not recommended because this would leave many numbers unused. Each animal might be assigned one number, which is then readable on the ear tag and also encoded in the RFID tag, but both

tags can also carry two independent numbers which then have to be linked in a data base.
Electronic identification of animals might also have additional advantages as it can be used for dairy or cattle and veal automation. However, the equipment currently used for these purposes such as neck belt transponders and the matching are not suitable to track individual animals. To control the feeding process, only ear tags are feasible, as the presence of the head in the feeding trough has to be registered. Individual identification can help to group the animals into weight classes or to link breeding information to weight outcome.
RFID technology is also widely used in logistics and the supply chain management and in many other business areas. In logistics, the Electronic Product Code (EPC) is used. The EPC is stored in the RFID transponder and consists of a header, which classifies the code type, and codes for the EPC manager (the GS1 member, who issued the code, e.g. the manufacturer), the object class (e.g. an article reference) and a serial number. Product codes are for example the SGTIN (Serialized Global Trade Item Number), the SSCC (Serial Shipping Container Code) or the GRAI (Global Returnable Asset Item). In the retail sector, RFID has never reached full coverage. The US supermarket and store corporation Wal-Mart required that their vendors place RFID tags on all shipments in 2005, but this rule seems to have never been totally enforced. One of the disadvantages of tagging individual items in the retail business are environmental concerns such as the consumption of resources (metals) and the difficulties for the recycling of tagged packaging.
The most commonly used code in the retail business is the European Article Number (EAN), also known as GTIN (Global Trade Item Number). The EAN standards are defined by GS1. The number has 13 or 8 digits (for smaller items). It is usually printed as a bar code on the sales packaging. Bar codes are more suitable for quick and safe automatic reading for example by scanner at a point of sale then human readable characters. The EAN 13 consists of a company prefix (7 to 9 digits) which includes the country prefix of the issuing GS1 member organisation and an individual item reference (3 to 5 digits, depending on the length of the prefix). The last digit is a single checksum digit. A licence fee has to be paid to obtain the company prefix from the GS1.

Fig. 3: Example for EAN 13 bar code.

Another bar code standard is the GS1-128 standard. It is a subset of the Code128 symbology used to code data into a bar code. The main use is in the logistics sector. Its maximum capacity is up to 48 characters, the maximum length of the bar code must not exceed 165 mm. The GS1-128 code contains a list of data fields, each identified by an two to four digit application identifier. Data fields of variable length must be terminated by a function code. The data given describe the product or the shipping such as the SSCC or the lot number etc. However, usually no individual serial number is coded in the GS1-128. For labelling of meat, the “Reference to Source Entity” can identify the individual ear tag number of the animal, the UN/ECE Meat Carcasses and cuts classification can

describe the product, the Country of Initial Processing gives the country of fattening etc. This standard is therefore suitable for tracking and tracing purposes in the meat industry.
Related to bar codes is the two dimensional DataMatrix code (GS1, 2010). It consists of black and white fields in a rectangular grid, which each code for one bit. The symbol also contains a L-shaped finder pattern, the code data itself and some error correction codewords. Each DataMatrix symbol can code for up to 2335 alphanumeric characters. The GS1 has defined a standard, then called GS1 DataMatrix. The Symbology Identifier (the first three characters transmitted by the scanner indicating symbology type) ']d2' specifies that the symbol read is a GS1 DataMatrix symbol. The symbol can be read as a concatenation of data elements such as GTIN, expiration date, serial number etc. with a fixed or variable length. Each data element string is identified by a two to four digit application identifier at its start position. To read a DataMatrix code, a camera based scanner has to be used, but even many modern mobile phones are capable of reading such a code. However, the GS1 DataMatrix is not intended for reading at a retail point of sale.

Fig. 4: Example for GS1 DataMatrix symbol

7. State-of-the-art in other industries

7.1. Product identification in the pharmaceutical industry

The purpose of product identification in the pharmaceutical industry is mainly the prevention of fraud and counterfeiting. To assure the origin of a product from a pharmaceutical manufacturer who guarantees for the safety of the ingredients, the production process and the final product, the smallest sales unit is marked with an identification code. This code is then recorded and can be used to document the position of the product in the supply chain using tracking and tracing procedures. The main goal is to ensure the consumer safety, but there are also billing and tax requirements.
Legal regulations and standards still differ on a national base. In Turkey, relevant legislation is currently being implemented. Each sales unit has to be marked with a unique serial number. The patient submits this package to the insurance company to receive a refund.
In California, the ePedrigree law ("Electronic California Prescription Drug Pedigree Law") obliges all pharmacies to sell prescription drugs only with a record documenting

every step in the supply chain from the manufacturer to the retailer. The law was originally planned to apply in 2011 but has been postponed to 2015.
In Italy, all pharmaceutical packages have to be individually labelled by prestamped labels ("Bollini").
The French CIP (Club Inter-Pharmaceutic) has recommended usage of a 13 character code for the product number starting in 2011. Individual serial numbers for each unit are not intended to be used.
Recommendations on product identification have been made by the IFAH (International Federation for Animal Health) and by the EFPIA (European Federation of Pharmaceutical Industries and Associations).
The data printed on the each package or other smallest sales unit, e.g. blister, bottle, etc. may contain the batch number, the GTIN (Global Trade Item Number), the date of production, the data of expiry, and a serial number. The serial number is unique and non-sequential. Data is usually coded into a 2D-DataMatrix code following the GS1 standard. The code is printed on the unit or a pre-printed label is applied. Each package is then optically inspected and the serial number is recorded in a database. During the packaging process from carton to pallet, a hierarchy of serial numbers is built. Finally all transactions within the supply chain down to the pharmacy and the end customer can be recorded in a database kept by the manufacturer.
In comparison to what can be achieved with product coding in more common retail products and the food sector, practices applied in the pharmacy sector allow for a much higher tracking and tracing precision due to unique serial numbering of single smallest sales unit. In the pharma sector, it is commonly easy to justify such an approach by the benefits created. In the food sector however, due to the lower value of the tracked goods high costs may not pay off.

7.2. Electronic record procedure for waste in Germany

In Germany, the management of all hazardous wastes has to be recorded on the way from producer to carrier and disposer. Since 1st of April 2010, these records have to be kept electronically. The goal is to enforce the product responsibility of the producer and to ensure a well-regulated disposal.
For all transactions, electronic documents are generated. Starting 1st of February 2011, these documents have to be electronically signed using a qualified electronic signature. This involves the use of a signature card for each person involved and a card reader. A central coordination site ("Zentrale Koordinierungsstelle Abfall") organizes the exchange of data between the waste management companies and the responsible authorities. The companies also exchange the documents between each other. Each company must also keep an electronic register of all transactions.
The interface for all data exchange is defined in XML schemas. Tag names in this XML dialect are in German to avoid misunderstandings of technical terms due to translation into English. While on a national level, this approach is feasible, it hints at the importance of multilingual electronic thesauri in international data exchange. The type of waste is recorded as given in the List of Wastes established by the European Commission. This catalogue classifies all waste types generated in the EU. It is a six digit code, where the first two digits determine the source such as a specific industry branch (e.g. 02 WASTES FROM AGRICULTURAL, HORTICULTURAL, HUNTING, FISHING AND AQUA-

CULTURAL PRIMARY PRODUCTION, FOOD PREPARATION AND PROCESSING), the next two a production process (e.g. 02 03 Wastes from fruit, vegetables, cereals, edible oils, cocoa, coffee and tobacco preparation and processing; tobacco processing; conserve production) and the last two digits the waste type (02 03 04 – materials unsuitable for consumption or processing).

Usually, for this electronic record procedure, a suitable software application is used. Various products are offered on the market. Web based access to the central coordination site is available free of charge. The electronic record procedure is meant to be applied only within Germany, export to other states is not covered. The reason for this restriction to national law is that all hazardous waste must be disposed in the same country where it was produced anyways.

8. Conclusions and Outlook

For each of the four building blocks identified (protocol, syntax, semantics, identification) for creating a backbone solution, several technologies exist to provide the necessary functionality.

Complex messaging protocols based on SOAP can be used in well-defined, controlled environments but will probably be too difficult to implement on a larger scale. RESTful web services have been proven to be better suited to networks with large numbers of small, anonymous stakeholders and a lack of control.

On the syntax level, XML is already widely used in the food and agricultural sector. It is thus well understood and can easily be implemented by most stakeholders. The disadvantage of XML is its inefficiency during data transfer on-the-wire due to its verbosity. This may result in problems on large scale tracking and tracing. There are however replacements for XML that are easy to handle and can be converted without much effort like e. g. JSON, so that syntax issues will not be a limiting factor in implementing the backbone.

Data items in the basic tracking and tracing data set described in deliverable 2.1 are semantically (relatively) well defined. The depicted backpack of additional data however is a very flexible and extensible container that calls for concise formalized and machine readable semantics of its content. A number of data dictionaries, thesauri and encoding systems exist in the food and agricultural sector that can be used for that purpose. Most of them however can not interoperate with other vocabularies at the moment or contain too few relationships between concepts and terms (i. e. are flat thesauri) to be useful for flexible and dynamic information exchange using e .g reasoning mechanisms to derive missing data. Thus increasing the usefulness of a European backbone solution by offering more than just simple tracking and tracing will reach its limits quickly if this problem is not tackled. The Food and Agricultural Organization of the United Nations (FAO) currently provides the KOS registry for collecting and referencing different knowledge organization and sharing systems in the agricultural and food sector thus providing a basis from which further harmonization and interconnection work could start.

Generating unique identifiers for objects is crucial in an open, global network. It usually requires hierarchical distribution of control over certain subsets of the value space. For identification in services, using URIs can be considered best practice. There are standard mechanisms for creating them and numbers can easily be turned into URIs by attaching a special prefix denoting the kind of identifier. Several organizations exist that can deal with maintenance of the upper levels of the hierarchy. The international domain name system registrars or GS1 offering company codes are examples. By careful consideration of the URI structure, compatibility with other numbering system can be achieved, even if not following their respective standard procedure of registration. As such, it would be possible to setup special identifier value spaces for use by small and medium sized enterprises that are not willing or not able to pay for taking part in more expensive identification systems.

To gain a clearer understanding of processes, a number of example tracking and tracing scenarios involving the special characteristics of the processes themselves and of the food products will be described and consequences, advantages and disadvantages of applying the technologies and methods mentioned above will be analyzed. From the experiences and knowledge gained, the backbone solution proposal specification can then be put together.

Consortium

Department for Food and Resource Economics, University of Bonn **(UBO)**	Germany
Kuratorium für Technik und Bauwesen in der Landwirtschaft e. V. **(KTBL)**	Germany
Department of Food Biotechnology and Food Process Engineering, Technische Universität Berlin **(TUB)**	Germany
The European Association for Food Safety **(SAFE)**	Belgium
The Swedish Institute for Food and Biotechnology AB **(SIK)**	Sweden
Faculty of Agricultural Sciences, University of Aarhus **(AU)**	Denmark
Centre for Food Policy, City University London **(City)**	United Kingdom
RLabs Market Research Ltd. **(RLabs)**	Greece
Faculty of Bio-engineering, Department of Agricultural Economics, Ghent University **(Ugent)**	Belgium
Campden BRI Magyarország Nonprofit Kft **(CCH)**	Hungary

Appendix A: References

T. Berners-Lee, R. Fielding, L. Masinter (1998): RFC 2396: Uniform Resource Identifiers (URI): Generic Syntax. Internet Engineering Task Force.

P. V. Biron, K. Permanente, A. Malhotra (2004): XML Schema Part 2: Datatypes Second Edition. World Wide Web Consortium.

J. Bosak, T. McGrath, G. K. Holman (2006): Universal Business Language v2.0. OASIS Universal Business Language (UBL) Technical Committee.

T. Bray, J. Paoli, C. M. Sperberg-McQueen, E. Maler, F. Yergeau (2008): Extensible Markup Language (XML) 1.0 (Fifth Edition). World Wide Web Consortium.

D. Brickley, R. V. Guha, B. McBride (2004): RDF Vocabulary Description Language 1.0: RDF Schema. World Wide Web Consortium.

R. Chinnici, J.-J. Moreau, A. Ryman, S. Weerawarana (2007a): Web Services Description Language (WSDL) Version 2.0 Part 1: Core Language. World Wide Web Consortium.

R. Chinnici, H. Haas, A. A. Lewis, J.-J. Moreau, D. Orchard, S. Weerawarana (2007b): Web Services Description Language (WSDL) Version 2.0 Part 2: Adjuncts. World Wide Web Consortium.

J. Clark, M. Makoto (2001): Relax NG Specification. Organization for the Advancement of Structured Information Standards (OASIS).

D. Crockford (2006): RFC 4627: The application/json Media Type for JavaScript Object Notation. Internet Engineering Task Force Network Working Group.

S. DeRose, E. Maler, D. Orchard, N. Walsh (2010): XML Linking Language (XLink) Version 1.1. World Wide Web Consortium.

J.-J. Dubray, S. St. Amand, M. J. Martin (2006): ebXML Business Process Specification Schema Technical Specification v2.0.4. ebXML Business Process Technical Committee.

EPCglobal Inc. (2007): EPC Information Services (EPCIS) Version 1.0.1 Specification.

EFSA (2010a): Standard sample description for food and feed. In: EFSA Journal 2010;8(1):1457. European Food Safety Authority, Parma, Italy.

EFSA (2010b): Guidance on Data Exchange. In: EFSA Journal 2010;8(11):1895. European Food Safety Authority, Parma, Italy.

D. C. Fallside, P. Walmsley (2004): XML Schema Part 0: Primer Second Edition. World Wide Web Consortium.

R. Fielding (2000): Architectural Styles and the Design of Network-based Software Architectures. University of California, Irvine.

S. Fuger, F. Najmi, N. Stojanovic (2005a): ebXML Registry Information Model Version 3.0. OASIS ebXML Registry Technical Committee.

S. Fuger, F. Najmi, N. Stojanovic (2005b): ebXML Registry Services and Protocols Version 3.0. OASIS ebXML Registry Technical Committee.

GS1 (2010): Introduction to GS1 DataMatrix – An introduction and technical overview of the most advanced GS1 Application Identifiers compliant symbology.

M. Gudgin, M. Hadley, N. Mendelsohn, J.-J. Moreau, H. F. Nielsen, A. Karmarkar, Y. Lafon (2007): SOAP Version 1.2 Part 1: Messaging Framework (Second Edition). World Wide Web Consortium.

M. Gudgin, M. Hadley, N. Mendelsohn, J.-J. Moreau, H. F. Nielsen, A. Karmarkar, Y. Lafon (2007): SOAP Version 1.2 Part 2: Adjuncts (Second Edition). World Wide Web Consortium.

P. Hayes, B. McBride (2004): RDF Semantics. World Wide Web Consortium.

G. Klyne, J. J. Carroll, B. McBride (2004): Resource Description Framework (RDF): Concepts and Abstract Syntax. World Wide Web Consortium.

F. Manola, E. Miller, B. McBride (2004): RDF Primer. World Wide Web Consortium.

A. Miles, S. Bechhofer (2009): SKOS Simple Knowledge Organization System Reference. World Wide Web Consortium.

N. Mitra, Y. Lafon (2007): SOAP Version 1.2 Part 0: Primer (Second Edition). World Wide Web Consortium.

A. Møller, J. Ireland (2009): LanguaL 2009 – The LanguaL Thesaurus. EuroFIR Technical Report D1.8.43.

OASIS ebXML Collaboration Protocol Profile and Agreement TC (2002): Collaboration-Protocol Profile and Agreement Specification Version 2.0. OASIS ebXML Collaboration Protocol Profile and Agreement Technical Committee.

L. Richardson, S. Ruby (2007): RESTful Web Services. O'Reilly, Sebastopol.

H. S. Thompson, D. Beech, M. Maloney, N. Mendelsohn (2004): XML Schema Part 1: Structures Second Edition. World Wide Web Consortium.

UN/CEFACT (2003): Core Components Technical Specification – Part 8 of the ebXML Framework, Version 2.01. United Nations Centre for Trade Facilitation and Electronic Business.

UNECE (2004): UNECE Standard: Bovine Meat Carcasses and Cuts. 2004 Edition. United Nations Publication ECE/TRADE/326. United Nations Economic Commission for Europe, Working Party on Agricultural Quality Standards.

UNECE (2009): UNECE Standard: Turkey Meat Carcases and Parts. 2009 Edition. United Nations Publication ECE/TRADE/358. United Nations Economic Commission for Europe, Working Party on Agricultural Quality Standards.

UNECE (2006): UNECE Standard: Llama/Alpaca Meat Carcases and Cuts. 2006 Edition. United Nations Publication ECE/TRADE/368. United Nations Economic Commission for Europe, Working Party on Agricultural Quality Standards.

P. Wenzel (2007): OASIS ebXML Messaging Services Version 3.0: Part 1, Core. OASIS ebXML Messaging Services Technical Committee.

K. Zyp (2010): A JSON Media Type for Describing the Structure and Meaning of JSON Documents. Internet Engineering Task Force Informational Internet Draft.